394 WE CLAIM:

395   1. (canceled) rewritten/re-presented in claim 5

396   2. (canceled) rewritten/re-presented in claim 6

397   3. (canceled) rewritten/re-presented in claim 7

398   4. (canceled) rewritten/re-presented in claim 8

399   5. (canceled) rewritten/re-presented in claim 9

400   6. (canceled) rewritten/re-presented in claim 10

401   7. (canceled) rewritten/re-presented in claim 11

402   8. (canceled) rewritten/re-presented in claim 12

403   9. (re-presented - formerly independent claim 5) A method of recognizing an observed

404   subsequence as being generated by one of a set of Hidden Markov Models (HMM),

405   characterized by:

406   • the fact that it searches the subsequence, Q, that offer the minimization of an

407   inverse confidence measure, over all possible matchings,

408   • where the inverse confidence measure is one of

409   1) the accumulated posterior, normalized with the length of the matched sub-

410   sequence X_b^e (aka. 'simple normalization')

411

$$\frac{-1}{e-b+1} \log P(Q|X_b^e)$$

21

412   2) partitioning the states in a HMM into phonems, having a function

413   Phonemes(Q) that returns the segmentation of a path Q in the HMM into

414   phonems, and computing one of:

415   2a) the worst average match in a phoneme, called 'real fitting',

416

$$\operatorname*{argmin}_{Q} \left(\max_{\mathcal{Q} \in Phonemes(Q)} \frac{\sum_{q^k \in \mathcal{Q}} - \log P(q^k | x_k)}{|\{k | q^k \in \mathcal{Q}\}|} \right)$$

417   2b) double normalization of the accumulated posterior over the number of

418   phonemes, J, and over the number of acoustic samples, $e_j - b_j + 1$, where

419   e_j is the time frame where Q enters phoneme j, and b_j is the exit time

420   frame from each phoneme, j,

421

$$\frac{-1}{J} \sum_{j=1}^{J} \left(\frac{1}{e_j - b_j + 1} \sum_{n=b_j}^{e_j} \log P(q_j^n | x_n) \right)$$

422   • and allows for the optional revaluation of the alternatives that offer the high-

423   est scores of a mentioned confidence measure on the basis of another confidence

424   measure,

425   • and when based on the confidence measure called 'simple normalization' uses a

426   method that applies Viterbi decoding for a HMM obtained by extending the initial

427   one with a filler state just after start and one just before the termination state,

428   and estimates the emission probability of the filler states in an iterative manner

429   as being equal to the inverse confidence measure in the previous iteration,

430   and where the emission probability in the filler states in the first iteration can be

431   initialized to any floating point number, but the iteration stops:

432   i at convergence yielding the estimation of a keyword's boundaries and score

433   as the obtained boundaries and score of non-filler states of the HMM,

434   ii when the confidence measure descends under a threshold value, T, estimating

435   only the keyword existence,

436   iii when the emission probability of filler states, ε_0 is initialized with T and is

437   reestimated, as value of ε_1 at the end of the first iteration, to be higher than

438   T deciding keyword inexistence,

439   • or for any of the three confidence measures: 'simple normalization', 'double nor-

440   malization' or 'real fitting', uses a beam-search-like algorithm that considers the

441   emission probability of the filler state as zero, computes progressively for each

442   pair of sample and state of HMM a set of possible alternatives paths to reach it,

443   the computation of this set is based on the sets of paths that lead to the states that

444   can be associated to the previous sample and extended with transitions allowed

445   by the analyzed HMM,

446   where this set can be reduced by using appropriate (safe) rules for the given

447   confidence measure, ensuring the correctness of the inference,

448   and where this set can be also reduced by using heuristics, for speeding up the

449   computation despite the risk of reducing the theoretical quality of the recognition,

450   heuristics of which a fast version stores only the best match,

451   and for all confidence measures one can prune the set of alternatives with safe rules

452   guaranteeing optimality, where:

453 • the 'simple normalization' confidence measure with beam-search is used with a

454 safe pruning that discards a path Q_1 given the existence of a path Q_2 whenever

455 $S_2 < S_1$ and $L_1 < L2$, where S_1 and L_1 respectively S_2 and L_2 are the minus of

456 the cumulated log of posteriors along the paths, and the lengths of the paths, for

457 the paths Q_1 respectively Q_2, and which can be optimized by sorting competing

458 paths based on their cost

459 • the 'double normalization' confidence measure on HMMs where no path skips any

460 phoneme is used with a safe pruning that discards a path Q_1 given the existence

461 of a path Q_2 whenever one of the following tests succeed:

462 (a) $l_2 \geq l_1$, $A > 0$, $B \leq 0$ and $L_c^2 A + L_c B + C \geq 0$

463 (b) $l_2 \geq l_1$, $A \geq 0$, $B \geq 0$ and $C \geq 0$

464 (c) $l_2 \geq l_1$, $A \leq 0$, $C \geq 0$ and $L^2 A + L B + C \geq 0$

465 (d) $l_2 \geq l_1$, $A = 0$, $B < 0$ and $L B + C \geq 0$

466 where we denote by a_1, p_1, l_1, respectively by a_2, p_2 and l_2 the confidence measure

467 for the previously visited phonemes, the posterior in the current phoneme and

468 the length in the current phoneme for the path Q_1, respectively the path Q_2,

469 and we also use the notations $A = a_1 - a_2$, $B = (a_1 - a_2)(l_1 + l_2) + p_1 - p_2$,

470 $C = (a_1 - a_2) l_1 l_2 + p_1 l_2 - p_2 l_1$, $L = L_{max} - \max\{l_1, l_2\}$, $L_c = -B/2A$ and L_{max} is

471 the maximum acceptable length for a phoneme,

472 • the 'double normalization' confidence measure on HMMs where some paths skip

473 phonemes is used with a safe pruning that discards a path Q_1 given the existence

474 of a path Q_2 whenever $l_2 \geq l_1$, $A \geq 0$, $p_1 \geq p_2$ respectively $F_2 \geq F_1$,

24

475   where F_1 respectively F_2 are the number of visited phonemes for paths Q_1 and

476   Q_2,

477   • the 'real fitting' is used with the safe pruning: Q_2 is discarded in favor of another

478   path Q_1 if the confidence measure of the Real Fitting for the previous phonemes

479   is inferior (higher in value) for Q_2 compared with Q_1, and if $p_1 \leq p_2$ and $l_2 \leq l_1$,

480   where p_1, l_1, respectively p_2, l_2 represent the minus of the logarithm of the cumu-

481   lated posterior respectively the number of frames in the current phoneme for the

482   path Q_1 respectively Q_2,

483   • and besides the previously mentioned safe pruning, heuristic prunings are also

484   used for removing paths when $p > L_{max}P_{max}$ in any state or when $\frac{p}{l} > P_{max}$ at

485   the output from a phoneme,

486   where p and l are the values in the current phoneme for the minus of the logarithm

487   of cumulated posterior and for the length of the path that is discarded.

488   10. (re-presented - formerly dependent claim 6) The method of claim 9, where the method

489   is used to estimate the existence of keywords and their position in utterances, using

490   Hidden Markov Models that model keywords.

491   11. (re-presented - formerly dependent claim 7) The method of claim 9, where the method

492   is used to estimate the existence of biomolecular subsequences and their position in the

493   chains of DNA using hidden Markov models to model the searched subsequences, and

494   where these models can be obtained by trivial translation from generalized profiles.

495   12. (re-presented - formerly dependent claim 8) The method of claim 9, where it carries out

496   the estimation of the existence of objects and their position in images, characterized

497   by the fact that

498   • it uses models of objects as subsequences represented by Hidden Markov Models,

499   • namely sections through views of objects are modeled by Hidden Markov Models,

500   • it uses emission probabilities based on a distance computed between colors, sim-

501   ple distances being yield by a Gaussian with median at the target color, or a

502   normalized inverse of the Euclidean distance in the RGB space,

503   • wherein the Hidden Markov Models that model the objects can be structured of

504   distinct regions, that play in the frame of the method the role of the phonemes

505   in claim 9,

506   • and wherein the models of the objects can be modified in a dynamic manner during

507   decoding with respect to the transition properties (existence and probability) on

508   the basis of the so far accumulated information in the process.